The Milstein scheme for stochastic delay differential equations without using anticipative calculus*

P.E. Kloeden† T. Shardlow‡

November 24, 2010

Abstract

The Milstein scheme is the simplest nontrivial numerical scheme for stochastic differential equations with a strong order of convergence one. The scheme has been extended to the stochastic delay differential equations but the analysis of the convergence is technically complicated due to anticipative integrals in the remainder terms. This paper employs an elementary method to derive the Milstein scheme and its first order strong rate of convergence for stochastic delay differential equations.

35K90, 41A58, 65C30, 65M99, 60K35

Keywords: Taylor expansions, stochastic differential equations, delay equations, strong convergence, SDDE, Milstein method

1 Introduction

The Milstein scheme is the simplest nontrivial numerical scheme for stochastic ordinary differential equations that achieves a strong order of convergence higher than that of the Euler-Maruyama scheme. It was first derived by Milstein, who used the Itô formula to expand an integrand involving the solution in one of the error terms of the Euler-Maruyama scheme. The iterative repetition of this idea underlies the systematic derivation of stochastic Taylor expansions and numerical schemes of arbitrarily high strong and weak orders, as expounded in Kloeden & Platen [9], see also Milstein [12].

Consider the Itô Stochastic Differential Delay Equation (SDDE) on $\mathbf{R}^d$ on the time interval $[0, T]$ with delay $\tau > 0$ given by

$$dx(t) = f(x_t, t)\, dt + g(x_t, t)\, dw(t), \tag{1.1}$$

*Partially supported by the ARC (UK) and DAAD (Germany). The coauthors thanks the Isaac Newton Institute for Mathematical Sciences of the University of Cambridge, where they wrote part of this paper.

†Institut für Mathematik, Goethe-Universität Frankfurt, D-60325 Frankfurt a.M., Germany, kloeden@math.uni-frankfurt.de

‡School of Mathematics, University of Manchester M13 9PL, Manchester, UK. shardlow@maths.man.ac.uk. Partially supported by the ARC (UK) and DAAD (Germany)

subject to initial data $x_0 = \xi \in C([-\tau, 0], \mathbf{R}^d)$, where $w(t)$ is a $\mathbf{R}^m$ Brownian motion,

$$f\colon C([-\tau,0],\mathbf{R}^d)\times\mathbf{R}^+\to\mathbf{R}^d, \qquad g\colon C([-\tau,0],\mathbf{R}^d)\times\mathbf{R}^+\to\mathbf{R}^{d\times m},$$

and we use x_t to denote the segment $\{x(t+\theta)\colon \theta\in[-\tau,0]\}$. An analogue of the Milstein scheme has been derived and analysed for SDDEs by Hu *et al.* [4]. However, the proofs of convergence are technically complicated due to the presence of anticipative integrals in the remainder terms.

The main contribution of this paper is to provide an elementary method to derive the Milstein scheme for SDDEs that does not involve anticipative integrals and anticipative stochastic calculus. Following the approach used by Jentzen & Kloeden for random ordinary differential equations [7, 6] and stochastic partial differential equations [8], we use deterministic Taylor expansions of the coefficient functions, with lower order expansions being inserted into the right hand side of higher order ones to give a closed form for the expansion. (A similar idea, without the final insertion, was also considered in [5, 11]). The Itô formula is not used at all and our proofs are much simpler than in [4].

The paper is organised as follows. §2 introduces notations and §3 gives some theory for the SDDEs that we study. §4 derives the Milstein method (see (4.12)) by use of Taylor expansions and calculates the local truncation error by collecting the Taylor remainder terms. §5 estimates the size of the remainder terms and §6 gives the main convergence result in Theorem 6.1. In §7, we show the assumptions of Theorem 6.1 apply to the case of finitely many discrete delays. An appendix gives some extra details to the proofs.

2 Notation

Consider two Banach spaces X and Y with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$. Let $\mathcal{L}(X,Y)$ denote the set of bounded linear operators $L\colon X\to Y$ with the operator norm $\|L\|_{\mathrm{op}} = \sup_{0\neq x\in X}\|Lx\|_Y/\|x\|_X$. We consider the product space $X\times Y$ of X and Y as a Banach space with the norm $\|(x,y)\| = \|x\|_X + \|y\|_Y$ for $x\in X, y\in Y$. For a function $\phi\colon X\to Y$, denote by $D^j\phi$ the jth Frechet derivative. Let $C^n(X,Y)$ denote the space of functions from X to Y with n uniformly bounded Frechet derivatives. Following [1], we interpret the jth derivative of $\phi\in C^n(X,Y)$ as a member of $\mathcal{L}(X,\dots\mathcal{L}(X,Y))$ for $j=0,\dots,n$; then

$$\|D^j\phi(x)h_1\;\dots\;h_j)\|_Y \le \|D^j\phi\|_{\mathrm{op}}\|h_1\|_X\|h_2\|_X\dots\|h_j\|_X, \quad x,h_1,\dots,h_j\in X.$$

We use Taylor's theorem for $\phi\in C^2(X,Y)$, which for $x,x^*\in X$ says

$$\begin{aligned}\phi(x) =&\phi(x^*) + D\phi(x^*)(x-x^*)\\ &+\int_0^1 (1-h)D^2\phi(x^*+h(x-x^*))(x-x^*)^2\,dh.\end{aligned}\tag{2.1}$$

Let $(\Omega,\mathcal{F},\mathcal{F}_t,\mathbf{P})$ denote a standard filtered probability space. We assume the Brownian motion $w(t)$ is $\mathcal{F}_t$ adapted. We denote the expectation of a random variable X with respect to $\mathbf{P}$ by $\mathbf{E}X$ and conditional expectation with respect to the σ-algebra $\mathcal{F}_s$ by $\mathbf{E}[X|\mathcal{F}_s]$. Consider $\mathbf{R}^d$ with the norm $\|x\|_{\mathbf{R}^d} = \langle x, x\rangle^{1/2}$ for $x \in \mathbf{R}^d$. $L^2(\Omega, \mathbf{R}^d)$ is the space of square integrable random variables X taking values in $\mathbf{R}^d$ with norm $\mathbf{E}\Big[\|X\|^2_{\mathbf{R}^d}\Big]^{1/2}$.

We often use the Banach space $C([-\tau, 0], \mathbf{R}^d)$ of continuous functions $\eta\colon [-\tau, 0] \to \mathbf{R}^d$ for $\tau > 0$ with norm

$$\|\eta\|_\infty = \sup_{-\tau\le\theta\le 0} \|\eta(\theta)\|_{\mathbf{R}^d}.$$

We make use of the following inequalities: for any $a_1, a_2, \dots, a_N$ and $p \ge 2$,

$$\Big(\sum_{i=1}^N a_i\Big)^p \le N^{p-1} \sum_{i=1}^N a_i^p. \tag{2.2}$$

For any $\mathbf{R}^{d\times m}$ valued adapted process $z(s)$, there exists K_p such that

$$\mathbf{E}\Big[\sup_{0\le t\le T} \Big\| \int_0^t z(s)\, dw(s) \Big\|^p_{\mathbf{R}^d}\Big] \le K_p \Big(\int_0^T \mathbf{E}\|z(s)\|_F^2\, ds\Big)^{p/2}, \tag{2.3}$$

where $\|\cdot\|_F$ denotes the Frobenius norm (Burkholder-Gundy-Davis inequality).

Throughout the paper, K is a generic constant that varies from one place to another and depends on f, g, the initial data ξ, the interval of integration $[0, T]$, but is independent of the discretisation parameter and choice of time points $s, t \in [0, T]$. The notation $\mathcal{O}(n)$ is used to denote a quantity bounded by Kn.

3 Regularity of solution

Consider the Ito SDDE

$$dx(t) = f(x_t, t)\, dt + g(x_t, t)\, dw(t), \tag{3.1}$$

subject to initial data $x_s = \eta \in C([-\tau, 0], \mathbf{R}^d)$. We denote the solution by $x(t; s, \eta)$ for $t \ge s$ and the corresponding segment by $x_t(s, \eta) \in C([-\tau, 0], \mathbf{R}^d)$. Notice that $x(t) = x(t; 0, \xi)$ and $x_t = x_t(0, \xi)$ gives the solution of (1.1). We make the following assumption of f and g.

Assumption 3.1. *Suppose that* $f \in C^3(C([-\tau, 0], \mathbf{R}^d) \times \mathbf{R}^+, \mathbf{R}^d)$ *and* $g \in C^3(C([-\tau, 0], \mathbf{R}^d) \times \mathbf{R}^+, \mathbf{R}^{d\times m})$.

Under this condition, there exists a unique solution to (3.1); that is, there exists a unique continuous $\mathcal{F}_t$ adapted $\mathbf{R}^d$ valued process $x(t)$ such that $x_s = \eta$ and almost surely

$$x(t; s, \eta) = x(s) + \int_s^t f(x_r(s, \eta), r)\, dr + \int_s^t g(x_r(s, \eta), r)\, dw(r), \qquad t \ge s, \tag{3.2}$$

where the stochastic integral is interpreted in the Ito sense. Assumption 3.1 can be replaced by much weaker conditions without loosing existence and uniqueness; see [10, 13] for further background on SDDEs.

We will make use of the following results concerning the regularity of $x_t(s,\eta)$ and the derivative of $x(t;s,\eta)$ with respect to the initial condition η.

Theorem 3.2. *Suppose that $p \geq 2$. There exists $K > 0$ such that*

$$\mathbf{E}\Big[\sup_{s-\tau\leq t\leq T}\|x(t;s,\eta)\|^p_{\mathbf{R}^d}\Big] \leq K(1+\mathbf{E}\|\eta\|^p_\infty) \tag{3.3}$$

$$\mathbf{E}\Big[\sup_{s-\tau\leq t\leq T}\|x(t;s,\eta_1)-x(t;s,\eta_2)\|^p_{\mathbf{R}^d}\Big] \leq K\mathbf{E}\|\eta_1-\eta_2\|^p_\infty \tag{3.4}$$

for any $\mathcal{F}_s$ measurable $C([-\tau,0],\mathbf{R}^d)$ valued random variables η,η_1,η.

Proof. (3.3) is given by [13, p.152]. (3.4) follows by applying Gronwall inequality and (2.3) as in [15]. □

We will use the following assumption of the initial condition.

Assumption 3.3. *Suppose that the initial function $\xi \in C([-\tau,0],\mathbf{R}^d)$ is uniformly Lipschitz continuous from $[-\tau,0]$ to $\mathbf{R}^d$.*

Corollary 3.4. *Suppose that $p \geq 2$ and that the initial function ξ satisfies Assumption 3.3. There exists $K > 0$ such that*

$$\mathbf{E}\Big[\sup_{-\tau\leq s\leq t\leq T}\|x(t)-x(s)\|^p_{\mathbf{R}^d}\Big] \leq K|t-s|^{p/2} \tag{3.5}$$

where $x(t)$ is the solution of (1.1).

Proof. Consider $0 \leq s \leq t \leq T$. The integral form (3.2) implies that

$$x(t)-x(s) = \int_s^t f(x_q,q)\,dq + \int_s^t g(x_q,q)\,dw(q), \quad a.s.$$

As f and g are bounded functions, we see by applying (2.3) that

$$\mathbf{E}\Big[\sup_{0\leq s\leq t\leq T}\|x(s)-x(t)\|^p_{\mathbf{R}^d}\Big] \leq K|t-s|^{p/2}. \tag{3.6}$$

As ξ is Lipschitz, this is easily extended to give (3.5).

□

Theorem 3.5 (derivative in initial condition)**.** *Suppose that $0 \leq s \leq t \leq T$. For $h \in C([-\tau,0],\mathbf{R}^d)$, let y^h_t denote the solution to*

$$dy^h(t) = D_1 f(x_t(s,\eta),t)y^h_t\,dt + D_1 g(x_t(s,\eta),t)y^h_t\,dw(t), \qquad y^h_s = h \tag{3.7}$$

where D_1 denotes the derivative with respect to the first argument. Then $y^h(t)$ is the L^2 Frechet derivative of $x(t;s,\eta)$ with respect to η and

$$\sup_{\|h\|_\infty<1}\mathbf{E}\Big[\sup_{s\leq t\leq T}\Big\|\frac{x(t;s,\eta+\epsilon h)-x(t;s,\eta)}{\epsilon}-y^h(t)\Big\|^2_{\mathbf{R}^d}\Big] \to 0 \quad as\ \epsilon\to 0. \tag{3.8}$$

We write $\frac{\partial x(t;s,\eta)}{\partial \eta}h$ *for* $y^h(t)$ *and note that* $\frac{\partial x(t;s,\eta)}{\partial \eta} \in \mathcal{L}(C([-\tau,0],\mathbf{R}^d), L^2(\Omega,\mathbf{R}^d))$. *There exists* $K>0$ *such that, for* $\eta \in C([-\tau,0],\mathbf{R}^d)$,

$$\mathbf{E}\Big[\sup_{0\le s\le t\le T}\Big\|\frac{\partial x(t;s,\eta)}{\partial \eta}\Big\|_{\mathrm{op}}^2\Big] \le K. \tag{3.9}$$

Proof. (3.7)–(3.8) are derived in [15]. Because Df and Dg are bounded, equation (3.9) follows by applying standard techniques to (3.7). □

Theorem 3.6 (second derivative in initial condition)**.** *Under the assumptions of Theorem 3.5, there exists a second* L^2 *Frechet derivative for* $x(t;s,\eta)$, *which we denote by* $\frac{\partial^2}{\partial \eta^2}x(t;s,\eta)$, *in the space*

$$\mathcal{L}(C([-\tau,0],\mathbf{R}^d),\mathcal{L}(C([-\tau,0],\mathbf{R}^d),L^2(\Omega,\mathbf{R}^d))).$$

There exists $K>0$ *such that, for* $\eta \in C([-\tau,0],\mathbf{R}^d)$,

$$\mathbf{E}\Big[\sup_{0\le s\le t\le T}\Big\|\frac{\partial^2 x(t;s,\eta)}{\partial \eta^2}\Big\|_{\mathrm{op}}^2\Big] \le K. \tag{3.10}$$

Proof. By using the second variational equation in place of (3.7), this is similar to Theorem 3.5. □

Following §2, we consider the N times product space $C([-\tau,0],\mathbf{R}^d)^N$ as a Banach space with norm

$$\|\underline{\eta}\| = \|\eta_1\|_\infty + \|\eta_2\|_\infty + \cdots + \|\eta_N\|_\infty, \tag{3.11}$$

for $\underline{\eta}=(\eta_1,\eta_2,\dots,\eta_N)\in C([-\tau,0],\mathbf{R}^d)^N$. The product space $C([-\tau,0],\mathbf{R}^d)^N\times\mathbf{R}^M$ has the norm $\|(\underline{\eta},\underline{y})\| = \|\underline{\eta}\| + \|\underline{y}\|_{\mathbf{R}^M}$ for each $\underline{\eta}\in C([-\tau,0],\mathbf{R}^d)^N$ and $\underline{y}\in\mathbf{R}^M$.

Corollary 3.7. *Fix* $J\in\mathbf{N}$ *and let* X *be the product space* $C([-\tau,0],\mathbf{R}^d)^4\times\mathbf{R}^{4dJ}$. *Consider* $t_1,\dots,t_J\ge s$ *and* $\Psi\in C^2(X,\mathbf{R}^{m\times m})$. *For* $\underline{\eta}=[\eta_1,\dots,\eta_4]\in C([-\tau,0],\mathbf{R}^d)^4$, *let*

$$E(\underline{\eta}) = \mathbf{E}\Psi(\eta_1,\dots,\eta_4,x(t_1;s,\eta_1),\dots,x(t_J;s,\eta_4)).$$

Then, we can find $K>0$ *such that*

$$\|DE(\underline{\eta})\|_{\mathrm{op}}\le K,\qquad \|D^2E(\underline{\eta})\|_{\mathrm{op}}\le K,$$

for any $\mathcal{F}_s$ *measurable* $C([-\tau,0],\mathbf{R}^d)^4$ *valued random variables* $\underline{\eta}$.

Proof. We consider a simple case: let $t\ge s\ge 0$ and $\Psi\in C^2(C([-\tau,0],\mathbf{R}^d)\times\mathbf{R}^d,\mathbf{R})$ and $E(\eta)=\mathbf{E}\Psi(\eta,x(t;s,\eta))$ for $\eta\in C([-\tau,0],\mathbf{R}^d)$. For $h_1,h_2\in C([-\tau,0],\mathbf{R}^d)$,

$$\begin{aligned} DE(\eta)(h_1) =&\mathbf{E}\Big[D\Psi(\eta,x(t;s,\eta))(h_1,\frac{\partial x(t;s,\eta)}{\partial\eta}h_1)\Big]\\ D^2E(\eta)(h_1,h_2) =&\mathbf{E}\Big[D^2\Psi(\eta,x(t;s,\eta))\Big((h_1,\frac{\partial x(t;s,\eta)}{\partial\eta}h_1),(h_2,\frac{\partial x(t;s,\eta)}{\partial\eta}h_2)\Big)\Big]\\ &+\mathbf{E}\Big[D\Psi(\eta,x(t;s,\eta))\Big(0,\frac{\partial^2 x(t;s,\eta)}{\partial\eta^2}(h_1,h_2)\Big)\Big]. \end{aligned}$$

Consequently,

$$
\begin{aligned}
\|D^2E(\eta)\|_{\rm op} \le & \mathbf{E}\Big[\|D^2\Psi(x(t;s,\eta))\|_{\rm op}\Big(1+\Big\|\frac{\partial x(t;s,\eta)}{\partial\eta}\Big\|_{\rm op}\Big)^2\Big] \\
& + \mathbf{E}\Big[\|D\Psi(x(t;s,\eta))\|_{\rm op}\Big\|\frac{\partial^2 x(t;s,\eta)}{\partial\eta^2}\Big\|_{\rm op}\Big].
\end{aligned}
$$

As $\Psi \in C^2(C([-\tau,0],\mathbf{R}^d),\mathbf{R})$, we can find a constant K such that

$$
\|D^2E(\eta)\|_{\rm op} \le K\mathbf{E}\Big[\Big(1+\Big\|\frac{\partial x(t;s,\eta)}{\partial\eta}\Big\|_{\rm op}\Big)^2\Big] + K\mathbf{E}\Big[\Big\|\frac{\partial^2 x(t;s,\eta)}{\partial\eta^2}\Big\|_{\rm op}\Big],
$$

which is uniformly bounded (for any η) by (3.9) and (3.10). The argument for DB is similar.

For the general case, the main difference is the number of arguments in the function. Without giving details, (3.9) and (3.10) are easily applied to the derivative to give the proof in the general case. □

4 Derivation of the Milstein method and remainders

We now derive the Milstein method by using Taylor expansions. We treat f as a function from $C([-\tau,0],\mathbf{R}^d)\times\mathbf{R}$ to $\mathbf{R}^d$ and use the second order Taylor theorem for functions on a Banach space (in this case, we consider $C([-\tau,0],\mathbf{R}^d)\times\mathbf{R}$ as the Banach space with the norm $\|\eta\|_\infty + |t|$ for $(\eta,t)\in C([-\tau,0],\mathbf{R}^d)\times\mathbf{R}$). This approach gives $f(x_r,r)$ in terms of f and its Frechet derivatives evaluated at (x_s,s). A similar approach is taken to g. The derivation is simple compared to using an Ito formula on $f(x_s,s)$ and all the remainder terms in the Taylor expansion are written as non-anticipative integrals. The estimation of the size of the remainder terms is routine, and depends on the Ito isometry and Burkholder-Gundy-Davis inequality (2.3), and is given in the Appendix.

First, using (2.1), we write down the second order Taylor expansion for f and g. For $0\le s\le r$

$$
\begin{aligned}
& f(x_r,r) = f(x_s,s) + Df(x_s,s)(x_r-x_s,r-s) \\
& + \int_0^1 (1-h)D^2f\Big(x_s+h(x_r-x_s),s+h(r-s)\Big)\Big(x_r-x_s,r-s\Big)^2\,dh \\
& g(x_r,r) = g(x_s,s) + Dg(x_s,s)(x_r-x_s,r-s) \\
& + \int_0^1 (1-h)D^2g\Big(x_s+h(x_r-x_s),s+h(r-s)\Big)\Big(x_r-x_s,r-s\Big)^2\,dh.
\end{aligned}
\tag{4.1}
$$

We write

$$
f(x_r,r) = f(x_s,s) + R_f(r;s,x_s) \tag{4.2}
$$

$$
g(x_r,r) = g(x_s,s) + Dg(x_s,s)(x_r-x_s,r-s) + R_g(r;s,x_s), \tag{4.3}
$$

where the remainder terms R_f and R_g are defined by (4.1). Substitute the expansions for f and g into the integral form (3.2) on $[s,t]$, to find

$$\begin{aligned} x(t) =& x(s) + f(x_s,s)(t-s) + g(x_s,s)\int_s^t dw(r) \\ &+ \int_s^t Dg(x_s,s)(x_r - x_s, r-s)\,dw(r) + R_1(t;s,x_s), \qquad \text{a.s.} \end{aligned}$$

where the remainder $R_1(t;s,x_s)$ is defined by

$$R_1(t;s,x_s) = \int_s^t R_f(r;s,x_s)\,dr + \int_s^t R_g(r;s,x_s)\,dw(r). \tag{4.4}$$

Let $I(s,t) = \int_s^t dw(r) = w(t) - w(s)$. Then, writing $x(t) = x(t;s,x_s)$, we have for $s \geq 0$

$$\begin{aligned} x(t;s,x_s) =& x(s) + f(x_s,s)(t-s) + g(x_s,s)I(s,t) \\ &+ \int_s^t Dg(x_s,s)(x_r(s,x_s) - x_s, r-s)\,dw(r) + R_1(t;s,x_s) \qquad \text{a.s.} \end{aligned}$$

and

$$x(t;s,x_s) = x(s) + g(x_s,s)I(s,t) + R_2(t;s,x_s), \qquad \text{a.s.} \tag{4.5}$$

where $R_2(t;s,x_s)$ is given by

$$\begin{aligned} R_2(t;s,x_s) =& f(x_s,s)(t-s) \\ &+ \int_s^t Dg(x_s,s)\Big(x_r(s,x_s) - x_s, r-s\Big)\,dw(r) + R_1(t;s,x_s). \end{aligned}$$

For $\theta \in [-\tau, 0]$ and $s + \theta \leq 0 \leq t + \theta$,

$$x(t+\theta;s,x_s) = x(s+\theta) + \Big(x(t+\theta;0,x_0) - x(0)\Big) + \Big(x(0) - x(s+\theta)\Big)$$

because $x(t+\theta;0,x_0) = x(t+\theta;s,x_s)$. By (4.5) and $x_0 = \xi$

$$\begin{aligned} x(t+\theta;s,x_s) =& x(s+\theta) + \Big(g(x_0,0)I(0,t+\theta) + R_2(t+\theta;0,x_0)\Big) \\ &+ \Big(\xi(0) - \xi(s+\theta)\Big). \end{aligned} \tag{4.6}$$

For $t + \theta < 0$,

$$x(t+\theta;s,x_s) = \xi(t+\theta) = x_s(s+\theta) + (\xi(t+\theta) - \xi(s+\theta)). \tag{4.7}$$

We combine (4.5)–(4.7), to get an expression for the segment x_t as a perturbation of x_s for any $0 \leq s \leq t$.

$$x_t(s,x_s)(\theta) = x_s(\theta) + \begin{cases} (\text{see } (4.5)), & 0 \leq s+\theta, \\ (\text{see } (4.6)), & s+\theta \leq 0 \leq t+\theta, \\ (\text{see } (4.7)), & t+\theta < 0, \end{cases} \tag{4.8}$$

As $x_s(\theta) = x(s+\theta)$, the correction term for (4.5) depends on $\hat{x}_s \in C([-2\tau, 0], \mathbf{R}^d)$, which we define for $\theta \in [-2\tau, 0]$ by

$$\hat{x}_s(\theta) = \begin{cases} x(s+\theta), & 0 < s+\theta, \\ \xi(s+\theta), & -\tau < s+\theta \le 0, \\ \xi(-\tau), & s \le -\tau. \end{cases}$$

The choice of constant for $s \le -\tau$ ensures continuity. We further require the following notations:

1. $G\colon C([-2\tau, 0], \mathbf{R}^d) \times \mathbf{R}^+ \to \mathcal{L}(C([-\tau, 0], \mathbf{R}^d), C([-\tau, 0], \mathbf{R}^d))$, defined by

$$G(\zeta, s)\eta(\theta) = \begin{cases} g(\pi_\theta \zeta, s+\theta)\eta(\theta), & s+\theta > 0, \\ g(\xi, 0)\eta(\theta), & s+\theta \le 0, \end{cases}$$

where $\zeta \in C([-2\tau, 0], \mathbf{R}^d)$, $s \in \mathbf{R}^+$, $\eta \in C([-\tau, 0], \mathbf{R}^d)$, $\theta \in [-\tau, 0]$, and $\pi_\theta : C([-2\tau, 0], \mathbf{R}^d) \to C([-\tau, 0], \mathbf{R}^d)$ is defined by $\pi_\theta \zeta(\phi) = \zeta(\phi + \theta)$ for $\phi \in [-\tau, 0]$.

2. define $I_t(s) \in C([-\tau, 0], \mathbf{R}^d)$ by

$$I_t(s)(\theta) = \begin{cases} I(s+\theta, t+\theta), & 0 \le s+\theta \le t+\theta, \\ I(0, t+\theta), & -\tau \le s+\theta \le 0 \le t+\theta, \\ 0, & \text{otherwise.} \end{cases}$$

3. For $s \le t$,

$$\delta_t(s)(\theta) = \begin{cases} 0, & 0 \le s+\theta, \\ \xi(0) - \xi(s+\theta), & s+\theta \le 0 \le t+\theta, \\ \xi(t+\theta) - \xi(s+\theta), & t+\theta \le 0. \end{cases}$$

Using this notation, we have from (4.8) that

$$x_t(s, x_s) = x_s + G(\hat{x}_s, s) I_t(s) + \delta_t(s) + R_{2,t}(s, \hat{x}_s), \qquad \text{a.s.} \tag{4.9}$$

with $R_{2,t}(s, \hat{x}_s) \in C([-\tau, 0], \mathbf{R}^d)$ defined by

$$R_{2,t}(s, \hat{x}_s)(\theta) = \begin{cases} R_2(t+\theta; s+\theta, \pi_\theta \hat{x}_s), & 0 \le s+\theta, \\ R_2(t+\theta; 0, \xi), & s+\theta < 0 \le t+\theta, \\ 0, & t+\theta \le 0. \end{cases}$$

Let

$$R(t; s, x_s) = R_1(t; s, x_s) + \int_s^t Dg(x_s, s)(R_{2,r}(s, \hat{x}_s), 0)\, dw(r). \tag{4.10}$$

Then, almost surely,

$$\begin{aligned} x(t;s,x_s) = & x(s) + f(x_s,s)(t-s) + g(x_s,s)I(s,t) \\ & + \int_s^t Dg(x_s,s)\Big(G(\hat{x}_s,s)I_r(s) + \delta_r(s), r-s\Big)\,dw(r) + R(t;s,x_s). \end{aligned} \tag{4.11}$$

Let $N \in \mathbf{N}$ denote a discretisation parameter and define a time step $\Delta t = T/N$ and a grid of points $t_k = k\Delta t$ for $k = 0,\ldots,N$. We now introduce the process $x^{\Delta t}(t)$: let $x^{\Delta t}(0) = \xi$ and for $t_k < t \le t_{k+1}$ let $x^{\Delta t}(t)$ solve

$$\begin{aligned} x^{\Delta t}(t) = & x^{\Delta t}(t_k) + f(x^{\Delta t}_{t_k},t_k)(t-t_k) + g(x^{\Delta t}_{t_k},t_k)\int_{t_k}^t dw(r) \\ & + \int_{t_k}^t Dg(x^{\Delta t}_{t_k},t_k)\Big(G(\widehat{x^{\Delta t}_{t_k}},t_k)I_r(t_k) + \delta_r(t_k), r-t_k\Big)\,dw(r). \end{aligned} \tag{4.12}$$

This is a continuous time process and at the grid points $t = t_k$, the evaluation of $x^{\Delta t}(t_k)$ is a simple update rule: $x^{\Delta t}(t_{k+1})$ depends on f and g evaluated at $(x^{\Delta t}(t_k), t_k)$ and on the random variables

$$\Delta w_k = \int_{t_k}^{t_{k+1}} dw(r) = w(t_{k+1}) - w(t_k), \quad Z_k = \int_{t_k}^{t_{k+1}} I_r(t_k)\,dw(r).$$

Note that $Z_k \in C([-\tau,0],\mathbf{R}^d)$ and

$$Z_k(\theta) = \begin{cases} \int_{t_k}^{t_{k+1}}\int_{t_k+\theta}^{r+\theta} dw(q)\,dw(r), & 0 \le t_k+\theta \le r+\theta. \\ \int_{t_k}^{t_{k+1}} w(r+\theta)\,dw(r), & -\tau \le t_k+\theta \le 0 \le r+\theta, \\ 0, & \text{otherwise.} \end{cases}$$

Given the initial data ξ, samples of Δw_k and Z_k, and the ability to evaluate f, g and Dg, $x^{\Delta t}(t_k)$ is easily computed and $x^{\Delta t}(t_k)$ is known as Milstein's method for approximating the solution of (1.1). Our main result (Theorem 6.1) gives conditions so that $x^{\Delta t}(t)$ is a first order approximation in the mean square sense to $x(t)$ over an interval $[0,T]$.

The practical implementation of this method can be more difficult. Δw_k are independent Gaussian random variables and are easily sampled. However, Z_k is a generalisation of the iterated stochastic integral and in general this problem is not well understood. See, for example, [9] for an approximate method for sampling Z_k.

5 Estimation of the remainder

The main result of this paper, given in Theorem 6.1, concerns the mean square convergence of the Milstein approximation $x^{\Delta t}(t)$ defined in (4.12). First, we estimate the size of the remainder term that was dropped when deriving the Milstein method (4.12) from (4.11).

Lemma 5.1. *Let Assumptions 3.1 and 3.3 hold. There exists $K > 0$ such that the remainder R defined in (4.10) satisfies*

$$\mathbf{E}\Big[\sup_{t_k \le t} \Big\| \sum_{j=0}^{k-1} R(t_{j+1}; t_j, x_{t_j}) \Big\|_{\mathbf{R}^d}^2 \Big] \le \sup_{t_k \le t} 2\mathbf{E}\Big[\sum_{i,j=0}^{k-1} \langle R_X(t_i), R_X(t_j) \rangle \Big] + K \Delta t^2, \tag{5.1}$$

for $0 \le t \le T$ and $\Delta t > 0$, where

$$R_X(t_i) = \int_{t_i}^{t_{i+1}} Df(x_{t_i}, t_i)(G(\hat{x}_{t_i}, t_i) I_r(t_i), 0)\, dr. \tag{5.2}$$

Proof. If $S_k = \sum_{j=0}^{k-1} r_{j+1}$, where r_k are $\mathbf{R}^d$ valued $\mathcal{F}_{t_k}$ measurable random variables, then $S_k - \mathbf{E}S_k$ is a discrete martingale and Doob's maximal inequality gives $\mathbf{E}\sup_{k \le n} \|S_k - \mathbf{E}S_k\|_{\mathbf{R}^d}^2 \le 2\mathbf{E}\|S_n - \mathbf{E}S_n\|_{\mathbf{R}^d}^2 \le 4\mathbf{E}\|S_n\|_{\mathbf{R}^d}^2 + 4\|\mathbf{E}S_n\|_{\mathbf{R}^d}^2$. Hence,

$$\begin{aligned}
\mathbf{E}\sup_{k \le n} \|S_k\|_{\mathbf{R}^d}^2 \le& 8\mathbf{E}\|S_n\|_{\mathbf{R}^d}^2 + 10 \sup_{k \le n} \|\mathbf{E}S_k\|_{\mathbf{R}^d}^2 \\
&\le 8\mathbf{E}\|S_n\|_{\mathbf{R}^d}^2 + 10 \sup_{k \le n} \mathbf{E}\|S_k\|_{\mathbf{R}^d}^2 \le 18 \sup_{k \le n} \mathbf{E}\|S_k\|_{\mathbf{R}^d}^2,
\end{aligned}$$

because $\|\mathbf{E}X\| \le \mathbf{E}\|X\| \le (\mathbf{E}\|X\|^2)^{1/2}$. Now

$$\begin{aligned}
\mathbf{E}\Big[\sup_{t_k \le t} \Big\| \sum_{j=0}^{k-1} R(t_{j+1}; t_j, x_{t_j}) \Big\|_{\mathbf{R}^d}^2 \Big] &\le 8 \sup_{t_k \le t} \mathbf{E}\Big[\Big\| \sum_{j=0}^{k-1} R(t_{j+1}; t_j, x_{t_j}) \Big\|_{\mathbf{R}^d}^2 \Big] \\
&+ 10 \sup_{t_k \le t} \Big\| \mathbf{E} \sum_{j=0}^{k-1} R(t_{j+1}; t_j, x_{t_j}) \Big\|_{\mathbf{R}^d}^2.
\end{aligned}$$

From (4.10), and the definition of R_1 in (4.4),

$$\begin{aligned}
R(t; s, x_s) = \int_s^t R_f(r; s, x_s)\, dr + \int_s^t R_g(r; s, x_s)\, dw(r) \\
+ \int_s^t Dg(x_s, s)(R_{2,r}(s, \hat{x}_s), 0)\, dw(r).
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbf{E}\Big[\Big\| \sum_{j=0}^{k-1} R(t; t_j, x_{t_j}) \Big\|_{\mathbf{R}^d}^2 \Big] &\le 3\mathbf{E} \Big\| \sum_{j=0}^{k-1} \int_{t_j}^{t_{j+1}} R_f(r; t_j, x_{t_j})\, dr \Big\|_{\mathbf{R}^d}^2 \\
&+3 \sum_{j=0}^{k-1} \int_{t_j}^{t_{j+1}} \mathbf{E}\|R_g(r; t_j, x_{t_j})\|_F^2\, dr \\
&+3 \sum_{j=0}^{k-1} \int_{t_j}^{t_{j+1}} \mathbf{E}\|Dg(x_s, s)(R_{2,r}(s, \hat{x}_s), 0)\|_F^2\, dr.
\end{aligned}$$

The last two term terms are estimated using standard Ito calculus techniques. We give the statements of the estimates in the Appendix (Lemmas A.1 and A.3) and find both terms are $\mathcal{O}(\Delta t^2)$ when estimated in the mean square sense. For the first term, we further develop R_f from (4.1).

$$
\begin{aligned}
&R_f(r;s,x_s)\\
&= Df(x_s,s)(x_r-x_s,r-s)\\
&+\int_0^1 (1-h)D^2 f\Big(x_s+h(x_r-x_s),s+h(r-s)\Big)\Big(x_r-x_s,r-s\Big)^2\,dh
\end{aligned}
$$

(substituting (4.9))

$$
\begin{aligned}
=&Df(x_s,s)(df(x_s,s)G(\hat{x}_s,s)I_r(s),r-s)+Df(x_s,s)(\delta_r(s)+R_{2,r}(s,\hat{x}_s),0)\\
&+\int_0^1 (1-h)D^2 f\Big(x_s+h(x_r-x_s),s+h(r-s)\Big)\Big(x_r-x_s,r-s\Big)^2\,dh
\end{aligned}
$$

(substituting (5.2))

$$
\begin{aligned}
=&R_X(s)+Df(x_s,s)(0,r-s)+Df(x_s,s)(\delta_r(s)+R_{2,r}(s,\hat{x}_s),0)\\
&+\int_0^1 (1-h)D^2 f\Big(x_s+h(x_r-x_s),s+h(r-s)\Big)\Big(x_r-x_s,r-s\Big)^2\,dh.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\Big\|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} R_f(r;t_j,x_{t_j})\,dr\Big\|_{\mathbf{R}^d}^2 = 2\Big\|\sum_{j=0}^{k-1} R_X(t_j)\Big\|_{\mathbf{R}^d}^2\\
&+2\Big\|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} Df(x_{t_j},t_j)(0,r-t_j)\\
&\quad + Df(x_{t_j},t_j)(\delta_r(t_j)+R_{2,r}(t_j,\hat{x}_{t_j}),0)+\int_0^1(1-h)\times\\
&\quad\times D^2 f\Big(x_{t_j}+h(x_r-x_{t_j}),t_j+h(r-t_j)\Big)\Big(x_r-x_{t_j},r-t_j\Big)^2\,dh\,dr\Big\|_{\mathbf{R}^d}^2.
\end{aligned}
$$

Note that the first term can be written

$$
2\Big\|\sum_{j=0}^{k-1} R_X(t_j)\Big\|_{\mathbf{R}^d}^2 = 2\sum_{i,j=0}^{k-1}\langle R_X(t_j),R_X(t_i)\rangle
$$

and this gives the first term in (5.1). The second term is $\mathcal{O}(\Delta t^2)$ in the sense of mean square. To see this, note

$$
\mathbf{E}\Big[\Big\|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} Df(x_{t_j},t_j)(0,r-t_j)\,dr\Big\|_{\mathbf{R}^d}^2\Big] \le K\Delta t^2
$$

as Df is bounded.

$$
\mathbf{E}\Big[\Big\|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} Df(x_{t_j},t_j)(\delta_r(t_j)+R_{2,r}(t_j,\hat{x}_{t_j}),0)\Big\|_{\mathbf{R}^d}^2\Big] \le K\Delta t^2
$$

as $\|\delta_r(t_j)\|_\infty \le K|t_j - r|$ because ξ is Lipschitz and Lemma A.3 controls $R_{2,r}$.

$$\mathbf{E}\Big[\Big\|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}}\int_0^1 (1-h)D^2 f\Big(x_{t_j}+h(x_r-x_{t_j}), t_j+h(r-t_j)\Big)\times \\ \times\Big(x_r-x_{t_j}, r-t_j\Big)^2 \,dh\,dr\Big\|_{\mathbf{R}^d}^2\Big] \le K\Delta t^2$$

by using the boundedness of $D^2 f$ and (3.5) (which applies because of Assumption 3.3). The proof is complete. $\square$

Unfortunately, the estimate in Lemma 5.1 is not always of size Δt^2. As Df and g (and hence G) are bounded, it follows from (5.2) that for a constant $K > 0$

$$\mathbf{E}\|R_X(t_i)\|_{\mathbf{R}^d}^2 \le \mathbf{E}\Big[\Big(\int_{t_i}^{t_{i+1}} \|Df\|_{\mathrm{op}}\|G(\hat{x}_{t_i}, t_i)\|_{\mathrm{op}}\|I_r(t_i)\|_\infty\,dr\Big)^2\Big] \le K\Delta t^3.$$

Hence, the sum of up to N^2 terms in (5.1) involving $R_X(t)$ gives a term of size $\mathcal{O}(\Delta t)$. We make the following assumption to gain an $\mathcal{O}(\Delta t^2)$ estimate.

Assumption 5.2. *For $\mathcal{P}_N = \{(i,j)\colon i,j = 0,\dots,N\}$, suppose there exists $\mathcal{Q}_N \subset \mathcal{P}_N$ such that for some $K > 0$*

$$\mathbf{E}\,\langle R_X(t_i), R_X(t_j)\rangle \le \begin{cases} K\Delta t^3, & (i,j)\in\mathcal{Q}_N \\ K\Delta t^4, & (i,j)\in\mathcal{P}_N - \mathcal{Q}_N \end{cases}, \quad \textit{for all } \Delta t > 0.$$

Clearly, the assumption holds with $\mathcal{Q}_N = \mathcal{P}_N$ and in the stochastic ordinary differential equation case with $\mathcal{Q}_N = \{(i,i)\colon i = 1,\dots,N\}$, as

$$R_X(t_i) = Df(x(t_i), t_i)\Big(g(x(t_i))\int_{t_i}^{t_{i+1}} w(r)\,dr, 0\Big)$$

and $\mathbf{E}\,\langle R_X(t_i), R_X(t_j)\rangle = 0$ for $i \ne j$ (for $i < j$, use the fact that $R_X(t_i)$ is $\mathcal{F}_{t_{i+1}}$ measurable and $\mathbf{E}\Big[R_X(t_j)|\mathcal{F}_{t_j}\Big] = 0$). We show in Theorem 7.4 that $\mathcal{Q}_N$ has $\mathcal{O}(N)$ members when f and g have finitely many discrete delays. Then, $\mathcal{O}(N)$ terms in (5.1) are of size Δt^3 and the remaining $\mathcal{O}(N^2)$ terms are of size Δt^4. This yields an $\mathcal{O}(\Delta t^2)$ estimate for (5.1).

6 Convergence theorem

We prove the following convergence theorem for Milstein's method.

Theorem 6.1. *Let Assumptions 3.1 and 3.3 hold. Let Assumption 5.2 hold when $\mathcal{Q}_N$ has $\mathcal{O}(N)$ members. For some constant $K > 0$, we have for any $\Delta t > 0$*

$$\Big(\mathbf{E}\sup_{t\in[-\tau,T]}\|x(t) - x^{\Delta t}(t)\|_{\mathbf{R}^d}^2\Big)^{1/2} \le K\Delta t. \tag{6.1}$$

Proof. From (4.11) and (4.12), we have almost surely for $t_k \le s < t_{k+1}$,

$$\begin{aligned}
x^{\Delta t}(s) - x(s) =& x^{\Delta t}(t_k) - x(t_k) \\
&+ \int_{t_k}^{\min\{t_{k+1},s\}} \Big(f(x_{t_k}^{\Delta t}, t_k) - f(x_{t_k}, t_k) \Big) \, dr \\
&+ \int_{t_k}^{\min\{t_{k+1},s\}} \Big(g(x_{t_k}^{\Delta t}, t_k) - g(x_{t_k}, t_k) \Big) dw(r) \\
&+ \int_{t_k}^{\min\{t_{k+1},s\}} \Big[Dg(x_{t_k}^{\Delta t}, t_k)(G(\widehat{x_{t_k}^{\Delta t}}, t_k) I_{r,t_k} + \delta_r(t_k), r - t_k) \\
&\qquad - Dg(x_{t_k}, t_k)(G(\widehat{x}_{t_k}, t_k) I_{r,t_k} + \delta_r(t_k), r - t_k) \Big] \, dw(r) \\
&+ R(s; t_k, x_{t_k}).
\end{aligned}$$

For $t_k \le s < t_{k+1}$, let

$$\begin{aligned}
D(s) =& f(x_{t_k}^{\Delta t}, t_k) - f(x_{t_k}, t_k) \\
M(s) =& \Big[g(x_{t_k}^{\Delta t}, t_k) - g(x_{t_k}, t_k) \Big] \\
&+ \Big[Dg(x_{t_k}^{\Delta t}, t_k)(G(\widehat{x_{t_k}^{\Delta t}}, t_k) I_{s,t_k} + \delta_s(t_k), s - t_k) \\
&\qquad - Dg(x_{t_k}, t_k)(G(\hat{x}_{t_k}, t_k) I_{s,t_k} + \delta_s(t_k), s - t_k) \Big].
\end{aligned}$$

Then, almost surely,

$$\begin{aligned}
x^{\Delta t}(s) - x(s) = \int_0^s D(r) \, dr + \int_0^s M(r) \, dw(r) \\
+ \sum_{j=0}^{k-1} R(t_{j+1}; t_j, x_{t_j}) + R(s, t_k, x_{t_k}).
\end{aligned}$$

For $t_k \le t < t_{k+1}$, let $e(t) = \mathbf{E} \sup_{s \le t} \|x^{\Delta t}(s) - x(s)\|_{\mathbf{R}^d}^2$.

$$\begin{aligned}
e(t) \le& 4\mathbf{E}\Big[\sup_{s \le t} \Big\| \int_0^s D(r) \, dr \Big\|_{\mathbf{R}^d}^2 \Big] + 4\mathbf{E}\Big[\sup_{s \le t} \Big\| \int_0^s M(r) \, dw(r) \Big\|_{\mathbf{R}^d}^2 \Big] \\
&+ 4\mathbf{E}\Big[\Big\| \sum_{j=0}^{k-1} R(t_{j+1}; t_j, x_{t_j}) \Big\|_{\mathbf{R}^d}^2 \Big] + 4\mathbf{E}\Big[\sup_{t_k \le s \le t} \|R(s; t_k, x_{t_k})\|_{\mathbf{R}^d}^2 \Big].
\end{aligned}$$

For $t_k \le t < t_{k+1}$,

$$\begin{aligned}
\mathbf{E}\Big[\sup_{s \le t} \Big\| \int_0^s D(r) \, dr \Big\|_{\mathbf{R}^d}^2 \Big] \le& K \int_0^t \mathbf{E}\Big[\sup_{s \le r} \|D(s)\|_{\mathbf{R}^d}^2 \Big] \, dr \\
\le& K \int_0^t \mathbf{E} \sup_{s \le r} \|f(x_s^{\Delta t}, s) - f(x_s, s)\|_{\mathbf{R}^d}^2 \, dr \\
\le& K \int_0^t e(r) \, dr.
\end{aligned}$$

By (2.3),

$$\mathbf{E}\Big[\sup_{s\le t}\Big\|\int_0^s M(r)\,dw(r)\Big\|^2_{\mathbf{R}^d}\Big] \le K\sup_{s\le t}\int_0^s \mathbf{E}\|M(r)\|_F^2\,dr.$$

Now, for $t_k \le s < t_{k+1}$,

$$\begin{aligned}\mathbf{E}\Big[\|M(s)\|_F^2\Big] \le& 2\mathbf{E}\Big[\|g(x^{\Delta t}_{t_k}, t_k) - g(x_{t_k}, t_k)\|_F^2\Big]\\ &+ 2\mathbf{E}\Big[\Big\|Dg(\widehat{x^{\Delta t}_{t_k}}, t_k)(G(\widehat{x^{\Delta t}_{t_k}}, t_k)I_{s,t_k} + \delta_s(t_k), s-t_k)\\ &\qquad - Dg(x_{t_k}, t_k)(G(\hat{x}_{t_k}, t_k)I_{s,t_k} + \delta_s(t_k), s-t_k)\Big\|_F^2\Big]\\ \le& K\,e(s),\end{aligned}$$

as $g \in C^3(C([-\tau, 0], \mathbf{R}^d), \mathbf{R}^{d\times m})$. Consequently,

$$\mathbf{E}\Big[\sup_{s\le t}\Big\|\int_0^s M(r)\,dw(r)\Big\|^2_{\mathbf{R}^d}\Big] \le K\int_0^t e(s)\,ds.$$

As $\mathcal{Q}_N$ has $\mathcal{O}(N)$ terms, Assumption 5.2 and Lemma 5.1 gives

$$\mathbf{E}\Big[\sup_{t_k\le t}\Big\|\sum_{j=0}^{k-1} R(t_{j+1}; t_j, x_{t_j})\Big\|^2_{\mathbf{R}^d}\Big] \le K\Delta t^2.$$

Putting the estimates together, we have

$$e(t) \le K\int_0^t e(s)\,ds + K\Delta t^2$$

and an application of Gronwall's inequality completes the proof. □

7 Discrete delays

We now give a particular class of SDDEs where Theorem 6.1 applies; namely, the class of SDDEs where f and g depend on a finite number of discrete delays. This case appears frequently in applications.

Assumption 7.1. *For $0 = \tau_1 < \tau_2 < \cdots < \tau_J \le \tau$, suppose that*

$$f(\eta, t) = F(\eta(-\tau_1), \dots, \eta(-\tau_J)), \qquad g(\eta, t) = G(\eta(-\tau_1), \dots, \eta(-\tau_J)),$$

for $\eta \in C([-\tau, 0], \mathbf{R}^d)$.

Lemma 7.2. *Suppose that Assumption 7.1 holds and $p \ge 2$. There exists $K > 0$ such that, for $0 \le r \le t \le T$ and $0 \le r \le s \le T$,*

$$\mathbf{E}\Big[\|\pi_s x_t(r,\eta)\|^p_\infty\Big] \le K|t-s|^{p/2}, \tag{7.1}$$

$$\mathbf{E}\Big[\|\pi_s x_t(r,\eta_1) - \pi_s x_t(r,\eta_2)\|^2_{\mathrm{op}}\Big] \le K|t-s|\,\|\eta_1-\eta_2\|^2_\infty, \tag{7.2}$$

where $\pi_s x_t(r,\eta) = x_t(r,\eta) - \mathbf{E}\Big[x_t(r,\eta)|\mathcal{F}_s\Big]$, for any $\eta, \eta_1, \eta_2 \in C([-\tau, 0], \mathbf{R}^d)$.

Proof. For $t+\theta \le s$, $\pi_s x_t(r,\eta)(\theta)=0$. For $t+\theta > s$ and $s \ge r$,

$$\begin{aligned}\pi_s x(t+\theta;r,\eta) =& x(t+\theta;r,\eta) - \mathbf{E}\Big[x(t+\theta;r,\eta)|\mathcal{F}_s\Big]\\ =& x(t+\theta;s,x_s(r,\eta)) - \mathbf{E}\Big[x(t+\theta;s,x_s(r,\eta))|\mathcal{F}_s\Big]\\ =& \int_s^{t+\theta} f(x_q(s,x_s(r,\eta)))\,dq + \int_s^{t+\theta} g(x_q(s,x_s(r,\eta)))\,dw(q)\\ &- \int_s^{t+\theta} \mathbf{E}\Big[f(x_q(s,x_s(r,\eta)))|\mathcal{F}_s\Big]\,dq\end{aligned}$$

and

$$\begin{aligned}\pi_s x(t+\theta;r,\eta) = &\int_s^{t+\theta} f(x_q(r,\eta)))\,dq + \int_s^{t+\theta} g(x_q(r,\eta))\,dw(q)\\ &- \int_s^{t+\theta} \mathbf{E}\Big[f(x_q(r,\eta))|\mathcal{F}_s\Big]\,dq.\end{aligned} \tag{7.3}$$

Now, inequality (2.3) gives the bound on $\mathbf{E}\|\pi_s x_t(r,\eta)\|_\infty^2$.

Denote $\partial_j f = \partial F(x_1,\dots,x_J)/\partial x_j \in \mathbf{R}^{d\times d}$. Under Assumption 7.1, $f(x_q(r,\eta))$ and $g(x_q(r,\eta))$ are mean square differentiable in η with uniformly bounded derivative. Hence,

$$\mathbf{E}\Big\|\frac{\partial f(x_q(r,\eta))}{\partial \eta}\Big\|_{\mathrm{op}}^p = \mathbf{E}\Big\|\sum_{j=1}^J \partial_j f(x_q(r,\eta))\frac{\partial x(q-\tau_j;r,\eta)}{\partial\eta}\Big\|_{\mathrm{op}}^p \le K$$

because $\|\partial x(q-\tau_j;r,\eta)/\partial\eta\|_{\mathrm{op}} \le 1$ for $q-\tau_j \le r$ and $\mathbf{E}\|\partial x(q-\tau_j;r,\eta)/\partial\eta\|_{\mathrm{op}}^2 \le K$ for $q-\tau_j \ge r$ from (3.9) of Theorem 3.5. A similar estimate holds for the partial derivatives in g. As the derivatives are uniformly bounded, the Lipschitz condition (7.2) is easily derived form (7.3). $\square$

Lemma 7.3. *Let Assumption 7.1 hold and let the initial data ξ satisfy Assumption 3.3. Suppose that $q+\Delta t \le r \le s$ and that Δ_q, Δ_s are mean zero $\mathbf{R}^m$ random variables such that*

$$\begin{aligned}&\Delta_q \text{ is } \mathcal{F}_{q+\Delta t} \text{ measurable and independent of } \mathcal{F}_q,\\ &\Delta_s \text{ is independent of } \mathcal{F}_s.\end{aligned}$$

Consider $h \in C^2(C([-\tau,0],\mathbf{R}^d)^4, \mathbf{R}^{m\times m})$. For some $K>0$,

$$\mathbf{E}\Big[\langle \Delta_q, h(x_q,x_r,x_s,x_{s+\Delta t})\Delta_s\rangle\Big] \le K\Delta t\Big(\mathbf{E}\|\Delta_q\|_{\mathbf{R}^m}^4 \mathbf{E}\|\Delta_s\|_{\mathbf{R}^m}^4\Big)^{1/4}, \qquad \Delta t > 0. \tag{7.4}$$

If $\Delta_q, \Delta_s \sim N(0,\sigma^2 I)$, then for some $K>0$

$$\mathbf{E}\Big[\langle \Delta_q, h(x_q,x_r,x_s,x_{s+\Delta t})\Delta_s\rangle\Big] \le K\Delta t\,\sigma^2.$$

Proof. Let $\underline{y} = [x_q, x_r, x_s, x_{s+\Delta t}]$ and let $\underline{y}^* = \mathbf{E}[\underline{y}|\mathcal{F}_s]$. Taylor's theorem provides

$$h(\underline{y}) = h(\underline{y}^*) + Dh(\underline{y}^*)\left(\underline{y} - \underline{y}^*\right) + R_h, \tag{7.5}$$

where the remainder satisfies $\|R_h\|_{\text{op}} \le \frac{1}{2}\|D^2 h\|_{\text{op}}\,\|\underline{y} - \underline{y}^*\|^2$. As the second derivative of h is uniformly bounded,

$$\|R_h\Delta_s\|_{\mathbf{R}^m} \le \|R_h\|_{\text{op}}\|\Delta_s\|_{\mathbf{R}^m} \le K\|\underline{y} - \underline{y}^*\|_\infty^2\|\Delta_s\|_{\mathbf{R}^m}.$$

By Lemma 7.2, for each $p \ge 2$, a constant K such that

$$\mathbf{E}\|\underline{y}^* - \underline{y}\|^p \le \mathbf{E}\|\pi_s y\|^p \le K\Delta t^{p/2} \tag{7.6}$$

(where the $\|\cdot\|$ is the product space norm defined in (3.11)). The Cauchy-Schwarz inequality gives

$$\begin{aligned}\mathbf{E}\Big[\|R_h\Delta_s\|_{\mathbf{R}^m}^2\Big] \le& K\Big(\mathbf{E}\Big[\|\underline{y} - \underline{y}^*\|_\infty^8\Big]\mathbf{E}\Big[\|\Delta_s\|_{\mathbf{R}^m}^4\Big]\Big)^{1/2}\\ \le& K\Delta t^2\mathbf{E}\Big[\|\Delta_s\|_{\mathbf{R}^m}^4\Big]^{1/2}.\end{aligned} \tag{7.7}$$

We now examine $h(\underline{y})\Delta_s$. Note that $\mathbf{E}\Big[\Delta_s|\mathcal{F}_{q+\Delta t}\Big] = 0$ and $h(\underline{y}^*)$ is $\mathcal{F}_s$ and hence is also $\mathcal{F}_{q+\Delta t}$ measurable. Therefore, $\mathbf{E}\Big[h(\underline{y}^*)\Delta_s|\mathcal{F}_{q+\Delta t}\Big] = 0$ and, using (7.5),

$$\mathbf{E}\Big[h(\underline{y})\Delta_s|\mathcal{F}_{q+\Delta t}\Big] = \mathbf{E}\Big[Dh(\underline{y}^*)(\underline{y} - \underline{y}^*)\Delta_s|\mathcal{F}_{q+\Delta t}\Big] + \mathbf{E}\Big[R_h\Delta_s|\mathcal{F}_{q+\Delta t}\Big].$$

We introduce the following notations. For $\underline{\eta} = [\eta_1, \eta_2] \in C([-\tau, 0], \mathbf{R}^d)^2$, let

$$\begin{aligned}\underline{Y}(\underline{\eta}) =& [\eta_1, x_r(q + \Delta t, \eta_2), x_s(q + \Delta t, \eta_2), x_{s+\Delta t}(q + \Delta t, \eta_2)]\\ \underline{Y}^*(\underline{\eta}) =& \mathbf{E}\Big[\underline{Y}(\underline{\eta})|\mathcal{F}_s\Big].\end{aligned}$$

For $\underline{\eta}, \underline{\nu} \in C([-\tau, 0], \mathbf{R}^d)^2$, by Lemma 7.2,

$$\mathbf{E}\Big[\|\underline{Y}(\underline{\eta}) - \underline{Y}^*(\underline{\eta})\|^p\Big] = \mathbf{E}\Big[\|\pi_s\underline{Y}(\underline{\eta})\|^p\Big] \le K\Delta t^{p/2} \tag{7.8}$$

$$\begin{aligned}\mathbf{E}\Big[\|\underline{Y}(\underline{\eta}) - \underline{Y}^*(\underline{\eta}) - \underline{Y}(\underline{\nu}) + \underline{Y}^*(\underline{\nu})\|^p\Big] =& \mathbf{E}\Big[\|\pi_s\underline{Y}(\underline{\eta}) - \pi_s\underline{Y}(\underline{\nu})\|^p\Big]\\ \le& K\Delta t^{p/2}\|\underline{\eta} - \underline{\nu}\|^p.\end{aligned} \tag{7.9}$$

By (3.4), $\mathbf{E}\|Y(\underline{\eta}) - Y(\underline{\nu})\|^p \le K\|\underline{\eta} - \underline{\nu}\|^p$ and hence

$$\mathbf{E}\Big[\|\underline{Y}^*(\eta) - \underline{Y}^*(\nu)\|^p\Big] \le K\|\underline{\eta} - \underline{\nu}\|^p. \tag{7.10}$$

Let $a(\underline{\eta}) = \mathbf{E}\Big[Dh(\underline{Y}^*(\underline{\eta}))\Big(\underline{Y}(\underline{\eta}) - \underline{Y}^*(\underline{\eta})\Big)\Delta_s\Big]$. In the case $\underline{\eta} = [x_q, x_{q+\Delta t}]$, we have $\underline{Y}(\underline{\eta}) = [x_q, x_r, x_s, x_{s+\Delta t}]$ and

$$\mathbf{E}\Big[h(\underline{y})\Delta_s|\mathcal{F}_{q+\Delta t}\Big] = a(\underline{\eta}) + \mathbf{E}\Big[R_h\Delta_s|\mathcal{F}_{q+\Delta t}\Big], \tag{7.11}$$

We derive a Lipschitz property for a. For any $\underline{\eta}, \underline{\nu} \in C([-\tau, 0], \mathbf{R}^d)^2$,

$$
\begin{aligned}
&\|a(\underline{\eta}) - a(\underline{\nu})\|_{\mathbf{R}^m} \\
\leq &\mathbf{E}\Big[\Big\| Dh(Y^*(\underline{\eta}))(\underline{Y}(\underline{\eta}) - \underline{Y}^*(\underline{\eta})) - Dh(\underline{Y}^*(\underline{\nu}))(\underline{Y}(\underline{\nu}) - \underline{Y}^*(\underline{\nu}))\Big\|_{\mathrm{op}}^2 \\
&\qquad\qquad \times \|\Delta_s\|_{\mathbf{R}^m}^2\Big]^{1/2} \\
\leq &\mathbf{E}\Big[\Big\| Dh(\underline{Y}^*(\underline{\eta}))(\underline{Y}(\underline{\eta}) - \underline{Y}^*(\underline{\eta})) - Dh(\underline{Y}^*(\underline{\nu}))(\underline{Y}(\underline{\nu}) - \underline{Y}^*(\underline{\nu}))\Big\|_{\mathrm{op}}^2\Big]^{1/2} \\
&\qquad \times \mathbf{E}\Big[\|\Delta_s\|_{\mathbf{R}^m}^2\Big]^{1/2}.
\end{aligned}
$$

As h has two bounded derivatives, Dh is Lipschitz and

$$
\begin{aligned}
&\Big\| Dh(\underline{Y}^*(\underline{\eta}))(\underline{Y}(\underline{\eta}) - \underline{Y}^*(\underline{\eta})) - Dh(\underline{Y}^*(\underline{\nu}))(\underline{Y}(\underline{\nu}) - \underline{Y}^*(\underline{\nu}))\Big\|_{\mathrm{op}} \\
\leq &\Big\|(Dh(\underline{Y}^*(\underline{\eta})) - Dh(\underline{Y}^*(\underline{\nu})))(\underline{Y}(\underline{\eta}) - \underline{Y}^*(\underline{\eta}))\Big\|_{\mathrm{op}} \\
&+ \Big\| Dh(\underline{Y}^*(\underline{\nu}))(\underline{Y}(\underline{\eta}) - \underline{Y}^*(\underline{\eta}) - \underline{Y}(\underline{\nu}) + \underline{Y}^*(\underline{\nu}))\Big\|_{\mathrm{op}} \\
\leq &K\Big\|\underline{Y}^*(\underline{\eta}) - \underline{Y}^*(\underline{\nu})\Big\| \; \Big\|\underline{Y}(\underline{\eta}) - \underline{Y}^*(\underline{\eta})\Big\| \\
&+ K\Big\|\underline{Y}(\underline{\eta}) - \underline{Y}^*(\underline{\eta}) - \underline{Y}(\underline{\nu}) + \underline{Y}^*(\underline{\nu})\Big\|.
\end{aligned}
$$

Hence, by (7.8)–(7.10),

$$
\begin{aligned}
&\mathbf{E}\Big[\Big\| Dh(\underline{Y}^*(\underline{\eta}))(\underline{Y}(\underline{\eta}) - \underline{Y}^*(\underline{\eta})) - Dh(\underline{Y}^*(\underline{\nu}))(\underline{Y}(\underline{\nu}) - \underline{Y}^*(\underline{\nu}))\Big\|_{\mathrm{op}}^2\Big] \\
\leq &K\mathbf{E}\Big[\|\underline{Y}^*(\underline{\eta}) - \underline{Y}^*(\underline{\nu})\|^2 \|\underline{Y}(\underline{\eta}) - \underline{Y}^*(\underline{\eta})\|^2\Big] + K\Delta t\|\underline{\eta} - \underline{\nu}\|^2 \\
\leq &K\mathbf{E}\Big[\|\underline{Y}^*(\underline{\eta}) - \underline{Y}^*(\underline{\nu})\|^4\Big]^{1/2} \mathbf{E}\Big[\|\underline{Y}(\underline{\eta}) - \underline{Y}^*(\underline{\eta})\|^4\Big]^{1/2} + K\Delta t\|\underline{\eta} - \underline{\nu}\|^2 \\
\leq &K\|\underline{\eta} - \underline{\nu}\|^2\Delta t + K\Delta t\|\underline{\eta} - \underline{\nu}\|^2.
\end{aligned}
$$

We conclude for some $K > 0$ that

$$
\|a(\underline{\eta}) - a(\underline{\nu})\|_{\mathbf{R}^m} \leq K\Delta t^{1/2}\mathbf{E}\Big[\|\Delta_s\|_{\mathbf{R}^d}^2\Big]^{1/2}\|\underline{\eta} - \underline{\nu}\|.
$$

Consider $\underline{\eta} = [x_q, x_{q+\Delta t}]$ and $\underline{\eta}^* = \Big[x_q, \mathbf{E}\Big[x_{q+\Delta t}|\mathcal{F}_q\Big]\Big]$. Δ_q has mean zero and is independent of $\mathcal{F}_q$ and therefore

$$
\mathbf{E}\Big[\big\langle \Delta_q, a(\underline{\eta})\big\rangle\Big] = \mathbf{E}\Big[\big\langle \Delta_q, a(\underline{\eta}) - a(\underline{\eta}^*)\big\rangle\Big].
$$

By Lemma 7.2, $\mathbf{E}\|\underline{\eta} - \underline{\eta}^*\|_\infty^p = \mathbf{E}\|\pi_q\underline{\eta}\|_\infty^p \leq K\Delta t^{p/2}$ and hence

$$
\begin{aligned}
\mathbf{E}\|a(\underline{\eta}) - a(\underline{\eta}^*)\|_{\mathbf{R}^m}^2 \leq &K\Delta t\mathbf{E}\Big[\|\Delta_s\|_{\mathbf{R}^m}^2\Big]\mathbf{E}\|\underline{\eta} - \underline{\eta}^*\|_\infty^2 \\
\leq &K\Delta t^2\mathbf{E}\Big[\|\Delta_s\|_{\mathbf{R}^m}^2\Big].
\end{aligned} \tag{7.12}
$$

To complete the proof of (7.4), note that

$$\mathbf{E}\Big[\,\langle \Delta_q, h(x_q, x_r, x_s, x_{s+\Delta t})\Delta_s\rangle\,\Big] = \mathbf{E}\Big[\,\Big\langle \Delta_q, \mathbf{E}\Big[h(\underline{y})\Delta_s|\mathcal{F}_{q+\Delta t}\Big]\Big\rangle\,\Big]$$

because Δ_q is $\mathcal{F}_{q+\Delta t}$ measurable. From (7.11),

$$\mathbf{E}\Big[\,\Big\langle \Delta_q, \mathbf{E}\Big[h(\underline{y})\Delta_s|\mathcal{F}_{q+\Delta t}\Big]\Big\rangle\,\Big] = \mathbf{E}\Big[\,\langle \Delta_q, a(\underline{\eta})\rangle + \Big\langle \Delta_q, \mathbf{E}\Big[R_h\Delta_s|\mathcal{F}_{q+\Delta t}\Big]\Big\rangle\,\Big].$$

We estimate the first term using (7.12)

$$\begin{aligned}\mathbf{E}\Big[\,\langle \Delta_q, a(\underline{\eta})\rangle\,\Big] \le&\mathbf{E}\Big[\|\Delta_q\|_{\mathbf{R}^m}\,\|a(\underline{\eta}) - a(\underline{\eta}^*)\|_{\mathbf{R}^m}\Big]\\ \le&\Big[\mathbf{E}\|\Delta_q\|^2_{\mathbf{R}^m}\,\mathbf{E}\Big[\|a(\underline{\eta}) - a(\underline{\eta}^*)\|^2_{\mathbf{R}^m}\Big]\Big]^{1/2}\\ \le&K\Delta t\mathbf{E}\Big[\|\Delta_q\|^2_{\mathbf{R}^m}\Big]^{1/2}\,\mathbf{E}\Big[\|\Delta_s\|^2_{\mathbf{R}^m}\Big]^{1/2}.\end{aligned}$$

Together with (7.7), this gives (7.4) as follows

$$\begin{aligned}&\mathbf{E}\Big[\,\Big\langle \Delta_q, \mathbf{E}\Big[h(\underline{y})\Delta_s|\mathcal{F}_{q+\Delta t}\Big]\Big\rangle\,\Big]\\ =&\mathbf{E}\Big[\,\langle \Delta_q, a(\underline{\eta})\rangle + \Big\langle \Delta_q, \mathbf{E}\Big[R_h\Delta_s|\mathcal{F}_{q+\Delta t}\Big]\Big\rangle\,\Big]\\ \le&K\Delta t\mathbf{E}\Big[\|\Delta_q\|^2_{\mathbf{R}^m}\Big]^{1/2}\,\mathbf{E}\Big[\|\Delta_s\|^2_{\mathbf{R}^m}\Big]^{1/2} + \Big(\mathbf{E}\Big[\|\Delta_q\|^2_{\mathbf{R}^m}\Big]\mathbf{E}\Big[\|R_h\Delta_s\|^2_{\mathbf{R}^m}\Big]\Big)^{1/2}\\ \le&K\Delta t\mathbf{E}\Big[\|\Delta_q\|^2_{\mathbf{R}^m}\Big]^{1/2}\,\mathbf{E}\Big[\|\Delta_s\|^2_{\mathbf{R}^m}\Big]^{1/2} + K\Delta t\Big(\mathbf{E}\Big[\|\Delta_q\|^2_{\mathbf{R}^m}\Big]\mathbf{E}\Big[\|\Delta_s\|^4_{\mathbf{R}^m}\Big]\Big)^{1/4}.\end{aligned}$$

The inequality (7.4) follows as $\big(\mathbf{E}\|\Delta_s\|^2_{\mathbf{R}^m}\big)^2 \le \mathbf{E}\|\Delta_s\|^4_{\mathbf{R}^m}$.

To complete the proof, recall that for mean zero Gaussian random variables the fourth moment is proportional to the second moment squared. □

The final theorem says that under the assumption of discrete delays, Assumption 7.1, we can show that $\mathcal{Q}_N$ has $\mathcal{O}(N)$ members and hence Theorem 6.1 applies and Milstein method converges in mean square with order one.

Theorem 7.4. *Suppose that Assumption 7.1 holds. If $\Delta t < \tau_2$ then Assumption 5.2 holds where $\mathcal{Q}_N$ has $\mathcal{O}(N)$ members. In particular, the error estimate* (6.1) *holds and the Milstein method converges with order one.*

Proof. Let $\Delta(s,t) = \int_s^t I(s,r)\,dr$. Then, R_X defined by (5.2) is given by

$$R_X(t_i) = \sum_{a=1}^{J} \partial_a f(x_{t_i}) g(x_{t_i-\tau_a}) \Delta(t_i - \tau_a, t_{i+1} - \tau_a).$$

Suppose that $t_i < t_j$.

$$\mathbf{E}\Big[\langle R_X(t_i), R_X(t_j)\rangle\Big] \tag{7.13}$$

$$= \sum_{a,b=1}^{J} \mathbf{E}\Big[\Big\langle \partial_a f(x_{t_i}) g(x_{t_i-\tau_a}) \Delta(t_i-\tau_a, t_{i+1}-\tau_a), \tag{7.14}$$

$$\partial_b f(x_{t_j}) g(x_{t_j-\tau_b}) \Delta(t_j-\tau_b, t_{j+1}-\tau_b)\Big\rangle\Big]$$

$$= \sum_{a,b=1}^{J} \mathbf{E}\Big[\Big\langle \Upsilon_{ij} \Delta(t_i-\tau_a, t_{i+1}-\tau_a), \Delta(t_j-\tau_b, t_{j+1}-\tau_b)\Big\rangle\Big], \tag{7.15}$$

where we define Υ_{ij} below in the following cases

1. for $t_i-\tau_a \le t_i \le t_{i+1}-\tau_a \le t_j-\tau_b \le t_j$
2. for $t_{i+1}-\tau_a \le t_i \le t_j-\tau_b \le t_j$
3. for $t_{i+1}-\tau_a \le t_j-\tau_b \le t_j$ and $t_i \ge t_j-\tau_b$
4. for $t_{j+1}-\tau_b \le t_i-\tau_a \le t_i < t_j$
5. $t_i-\tau_a \le t_{j+1}-\tau_b$ and $t_j-\tau_b \le t_{i+1}-\tau_a$.

We work out the contribution to $\mathbf{E}\Big[\langle R_X(t_i), R_X(t_j)\rangle\Big]$ for each case.

1. The condition $t_i \le t_{i+1}-\tau_a$ implies that $0 \le \Delta t-\tau_a$. As $\Delta t < \tau_2$, we must have $\tau_a=\tau_1=0$ and hence $t_i < t_{i+1} \le t_j-\tau_b \le t_j$. In this case, let

   $$\Upsilon_{ij} = \mathbf{E}\Big[g(x_{t_i})^T \partial_a f(x_{t_i})^T \partial_b f(x_{t_j}) g(x_{t_j-\tau_b}) | \mathcal{F}_{t_{j+1}-\tau_b}\Big],$$

   Using $x_{t_j} = x_{t_j}(t_{j+1}-\tau_b, x_{t_{j+1}-\tau_b})$ for $\tau_b \ge \Delta t$, we write

   $$\begin{aligned}\Upsilon_{ij} =& E_{a,b}(x_{t_i}, x_{t_j-\tau_b}, x_{t_j-\tau_b}, x_{t_{j+1}-\tau_b})\\ E_{a,b}(\eta_1,\eta_2,\eta_3,\eta_4) =& \mathbf{E}\Big[g(\eta_1)^T \partial_a f(\eta_1)^T \partial_b f(x_{t_j}(t_{j+1}-\tau_b, \eta_4)) g(\eta_3))\Big]\end{aligned}$$

   For $\tau_b=0$, we write

   $$E_{a,b}(\eta_1,\eta_2,\eta_3,\eta_4) = \mathbf{E}\Big[g(\eta_1)^T \partial_a f(\eta_1)^T \partial_b f(\eta_3) g(\eta_3)\Big]$$

   (the second argument is included to make h have four arguments).

   By Corollary 3.7, $E_{a,b} \in C^2(C([-\tau,0],\mathbf{R}^d)^4, \mathbf{R}^{m\times m})$ and Lemma 7.3 applies to h_{ab} with $p=t_i$, $r=t_j-\tau_b$, $s=t_j-\tau_b$, $\Delta_p = I(p,p+\Delta t)$ and $\Delta_s = I(s,s+\Delta t)$. Because Δ_p, Δ_s are $N(0,\sigma^2 I)$ with

   $$\sigma^2 = \int_0^{\Delta t}\int_0^{\Delta t} \mathbf{E}[w_1(r)w_1(s)]\,dr\,ds \le \Delta t^3,$$

   we find for some $K>0$ that

   $$\mathbf{E}\Big[\Big\langle \Upsilon_{ij}\Delta(t_i-\tau_a, t_{i+1}-\tau_a), \Delta(t_j-\tau_b, t_{j+1}-\tau_b)\Big\rangle\Big] \le K\Delta t^4. \tag{7.16}$$

2. for $t_{i+1}-\tau_a \le t_i \le t_j-\tau_b$,

$$\begin{aligned}\Upsilon_{ij} =& E_{a,b}(x_{t_i-\tau_a}, x_{t_i}, x_{t_j-\tau_b}, x_{t_{j+1}-\tau_b})\\ =&\mathbf{E}\Big[g(x_{t_i-\tau_a})^T \partial_a f(x_{t_i})^T \partial_b f(x_{t_j}) g(x_{t_j-\tau_b}) | \mathcal{F}_{t_{j+1}-\tau_b}\Big]\\ E_{a,b} =& \mathbf{E}\Big[g(\eta_1)^T \partial_a f(\eta_2)^T \partial_b f(x_{t_j}(t_{j+1}-\tau_b, \eta_4)) g(\eta_3)\Big]\end{aligned}$$

for $\tau_b \ge \Delta t$ and

$$E_{a,b} = \mathbf{E}\Big[g(\eta_1)^T \partial_a f(\eta_2)^T \partial_b f(\eta_3) g(\eta_3)\Big]$$

for $\tau_b = 0$. As in case 1., Lemma 7.3 applies with $p = t_i - \tau_a$, $r = t_i$, $s = t_j - \tau_b$.

3. for $t_{i+1}-\tau_a \le t_j-\tau_b \le t_j$ and $t_i \ge t_j - \tau_b$

$$\begin{aligned}\Upsilon_{ij} =& E_{a,b}(x_{t_i-\tau_a}, x_{t_{i+1}-\tau_a}, x_{t_j-\tau_b}, x_{t_{j+1}-\tau_b}),\\ =&\mathbf{E}\Big[g(x_{t_i-\tau_a})^T \partial_a f(x_{t_i})^T \partial_b f(x_{t_j}) g(x_{t_j-\tau_b}) | \mathcal{F}_{t_{j+1}-\tau_b}\Big]\\ E_{a,b} =& \mathbf{E}\Big[g(\eta_1)^T \partial_a f(x_{t_i}(x_{t_{j+1}-\tau_b}, \eta_4))^T \partial_b f(x_{t_j}(t_{j+1-\tau_b}, \eta_4)) g(\eta_3)\Big]\end{aligned}$$

for $\tau_b \ge \Delta t$ and

$$E_{a,b} = \mathbf{E}\Big[g(\eta_1)^T \partial_a f(x_{t_i}(x_{t_{j+1}-\tau_b}, \eta_4))^T \partial_b f(\eta_3) g(\eta_3)\Big]$$

for $\tau_b = 0$ (the second argument is included to make h have four arguments). Lemma 7.3 applies with $p = t_i - \tau_a$, $r = t_{i+1} - \tau_a$, $s = t_j - \tau_b$.

4. for $t_{j+1} - \tau_b \le t_i - \tau_a < t_i < t_j$

$$\begin{aligned}\Upsilon_{ij} =& E_{a,b}(x_{t_j-\tau_b}, x_{t_{j+1}-\tau_b}, x_{t_i-\tau_a}, x_{t_{i+1}-\tau_a})\\ =&\mathbf{E}\Big[g(x_{t_i-\tau_a})^T \partial_a f(x_{t_i})^T \partial_b f(x_{t_j}) g(x_{t_j-\tau_b}) | \mathcal{F}_{t_{i+1}-\tau_a}\Big].\\ E_{a,b} =& \mathbf{E}\Big[g(\eta_3)^T \partial_a f(x_{t_i}(t_{i+1}-\tau_a, \eta_4))^T \partial_b f(x_{t_j}(t_{i+1}-\tau_a, \eta_4)) g(\eta_1)\Big]\end{aligned}$$

for $\tau_a \ge \Delta t$ and

$$E_{a,b} = \mathbf{E}\Big[g(\eta_3)^T \partial_a f(\eta_3)^T \partial_b f(x_{t_j}(t_{i+1}-\tau_a, \eta_4)) g(\eta_1)\Big].$$

for $\tau_a = 0$. Lemma 7.3 applies with $p = t_j - \tau_b$, $r = t_{j+1} - \tau_b$, $s = t_i - \tau_a$.

5. This is the exceptional case and can occur for (a, b) only for $\mathcal{O}(N)$ pairs (i, j). The term Υ_{ij} is bounded (as it is defined in terms of f and g and their derivatives). It is easily to show

$$\mathbf{E}\Big[\Big\langle \Upsilon_{ij}\Delta(t_i-\tau_a, t_{i+1}-\tau_a), \Delta(t_j-\tau_b, t_{j+1}-\tau_b)\Big\rangle\Big] \le K\Delta t^3. \quad (7.17)$$

Let $\mathcal{Q}_N$ be the set of (i,j) such that condition 5. holds for at least one pair (a,b), where $a,b = 1,\ldots,J$. Notice that $\mathcal{Q}_N$ has $\mathcal{O}(N)$ members. For $(i,j) \in \mathcal{Q}_N$, we have a sum of $\mathcal{O}(J^2)$ terms of size $\mathcal{O}(\Delta t^3)$ (most of the terms are order $\mathcal{O}(\Delta t^4)$) and

$$\mathbf{E}\Big[\,\langle R_X(t_i), R_X(t_j)\rangle\,\Big] \le K\Delta t^3.$$

If $(i,j) \in \mathcal{P}_N - \mathcal{Q}_N$, no term in the sum (7.15) depends on condition 5., so that

$$\mathbf{E}\Big[\,\langle R_X(t_i), R_X(t_j)\rangle\,\Big] \le K\Delta t^4.$$

Hence $\mathcal{Q}_N$ has $\mathcal{O}(N)$ members and Theorem 6.1 completes the proof. $\square$

References

[1] A. Ambrosetti and G. Prodi, A Primer of Nonlinear Analysis, *Cambridge studies in advanced mathematics 34*, (1993), CUP.

[2] M. Arriojas, Y. Hu, S.-E. Mohammed and G. Pap, A delayed Black and Scholes formula, *Stoch. Anal. Appl.* **25** (2007), no. 2, 471–492.

[3] E. Buckwar, R. Kuske, S.-E. Mohammed and T. Shardlow, Weak convergence of the Euler scheme for stochastic differential delay equations. *LMS J. Comput. Math.* **11** (2008), 60–99.

[4] Y. Hu, S.-E. Mohammed and F. Yan, Discrete-time approximations of stochastic delay equations: the Milstein scheme. *Ann. Probab.* **32** (2004), no. 1A, 265–314.

[5] S. Janković and D. Ilić, An analytical approximate method for solving stochastic integro differential equations, *J. Math. Anal. Appl.*, **320** (2006), 230–245.

[6] P.E. Kloeden and A. Jentzen, Pathwise convergent higher order numerical schemes for random ordinary differential equations, *Proc. Roy. Soc. London A* **463** (2007), 2929–2944.

[7] A. Jentzen and P. E. Kloeden, Pathwise Taylor schemes for random ordinary differential equations, *BIT* **49** (1) (2009), 113–140.

[8] A. Jentzen and P.E. Kloeden, Taylor-expansions of solutions of stochastic partial differential equations with additive noise, *Annals of Probab.,* **38**, No. 2., (2010), 532–569.

[9] P.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Heidelberg, 1992; second revised printing 1999.

[10] X. Mao, *Stochastic Differential Equations and Applications*, Horwood, 1997.

[11] M. Milosević, M. Jovanović and S. Janković, An approximate method via Taylor series for stochastic functional differential equations, *J. Math. Anal. Appl.*, Vol. 363, (2010), 128–137.

[12] G.N. Milstein, *Numerical Integration of Stochastic Differential Equations*, Kluwer, Dordrecht, 1995.

[13] S.-E. Mohammed, Stochastic Functional Differential Equations, Research Notes in Mathematics No. 99, Pitman Books Ltd., London (1984).

[14] S.-E. Mohammed and T. Zhang, Anticipating stochastic differential systems with memory. *Stochastic Process. Appl.* **119** (2009), no. 9, 2773–2802.

[15] M. Reiss, Unpublished lecture notes, Humboldt University, 2010.

A Estimates

We derive some estimates on the remainder terms given in §4 on the time interval $[0, T]$.

Lemma A.1. *There exists $K > 0$ such that*

$$\mathbf{E}\Big[\sup_{0\le s\le r\le t\le T} \|R_f(r; s, x_s)\|^2_{\mathbf{R}^d}\Big] \le K|t-s|$$

$$\mathbf{E}\Big[\sup_{0\le s\le r\le t\le T} \|R_g(r; s, x_s)\|^2_{F}\Big] \le K|t-s|^2.$$

Proof. Using (3.3) and the regularity of f, g, both inequalities follow from the definitions in (4.2)–(4.3). □

Lemma A.2. *There exists $K > 0$ such that, for $0 \le s \le t \le T$,*

$$\mathbf{E}\Big[\sup_{s\le r\le t\le T} \|R_1(r; s, x_s)\|^2_{\mathbf{R}^d}\Big] \le K|t-s|^3.$$

Proof. Apply (2.2) and (2.3) to (4.4). Then,

$$\begin{aligned}&\mathbf{E}\Big[\sup_{s\le r\le t\le T} \|R_1(r; s, x_s)\|^2_{\mathbf{R}^d}\Big] \\ &\qquad \le 2\Big(\mathbf{E}\Big[\sup_{s\le r\le t\le T} \|R_f(r; s, x_s)\|^2_{\mathbf{R}^d}|t-s|^2\Big] + K\int_s^t \mathbf{E}\Big[\|R_g(r; s, x_r)\|^2_F\Big]\,dr\Big).\end{aligned}$$

The estimates in Lemma A.1 complete the proof. □

Lemma A.3. *There exists $K > 0$ such that, for $\Delta t > 0$,*

$$\mathbf{E}\Big[\sup_{s\le r\le s+\Delta t\le T} \|R_2(r; s, x_s)\|^2_{\mathbf{R}^d}\Big] \le K\Delta t^2$$

and

$$\mathbf{E} \sup_{s\le r\le s+\Delta t\le T} \|R_{2,r}(s, \hat{x}_s)\|^2_\infty \le K\Delta t^2. \tag{A.1}$$

Proof. By (2.2) and (2.3),

$$
\begin{aligned}
&\mathbf{E}\Big[\sup_{s\le r\le s+\Delta t\le T}\|R_2(r;s,x_s)\|^2_{\mathbf{R}^d}\Big]\\
\le&3\Big(K|t-s|^2+\hat{C}_2\sup_{s\le r\le s+\Delta t\le T}\int_s^r\mathbf{E}\Big[\|Dg(x_s,s)(x_p(s,x_s)-x_s,p-s)\|^2_F\Big]\,dp\\
&+\mathbf{E}\sup_{s\le r\le s+\Delta t\le T}\|R_1(r;s,x_s)\|^2_{\mathbf{R}^d}\Big)\\
\le&K\Big(\Delta t^2+\Delta t\Big(\mathbf{E}\Big[\sup_{s\le r\le s+\Delta t\le T}\|x_r(s,x_s)-x_s\|^2_\infty\Big]+\Delta t^2\Big)\\
&+\mathbf{E}\sup_{s\le r\le s+\Delta t\le T}\|R_1(r;s,x_s)\|^2_{\mathbf{R}^d}\Big).
\end{aligned}
$$

By Lemma A.1 and (3.5), this is $\mathcal{O}(\Delta t^2)$ as required. As $R_{2,r}$ is defined in terms of R_2 in (4.9), (A.1) follows easily. □